Aspect Based Sentiment Analysis on Students' Feedback Using Deep Learning

[1]Atharv Anand, [2]Akansha Naidu, [3]Sumit Kumar, [4]Satyajeet Kumar Sinha, [5]Supriya Bhosale
[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor
D Y Patil College of Engineering

Abstract - **Feedback from students is crucial for academic institutions to assess the performance of the faculty. Efficient handling of students' qualitative opinions while generating automatic report is a challenging task. Nonetheless, most institutions deal successfully with quantitative feedback, while qualitative feedback is either manually interpreted or ignored altogether. Therefore, we have proposed a supervised deep learning-based method which strives to improve the quality of feedback system by considering the sentiments along with their respective aspects. The proposed system is based on two-layer LSTM model which uses BERT for word embeddings. The pre-processed data is fed to the BERT embedding layer whose output is further passed on to the LSTM layers for aspect extraction and polarity classification. The task of aspect extraction is conducted by the LSTM layer1 and that of polarity classification by the LSTM layer 2.**

keywords **- Aspect Extraction, Sentiment Analysis, Deep Learning, LSTM, BERT, Student Feedback**

I. INTRODUCTION

Sentiment Analysis or Opinion Mining is the technique of comprehending the views or thoughts of a person towards an entity. It is basically a way of identifying and categorizing sentiments expressed in textual format to determine their attitude. The attitudes are commonly known as the polarity which is categorized as - positive, negative or neutral. Sentiment analysis is used in various sectors such as marketing, medical, education, etc. to improve their business and to help their consumers. Sentiment analysis is a treasure for human resource or service-based institutions as user feedback plays a crucial role in improving the quality and approach of service or product. Educational institutions like schools and colleges conduct sentiment analysis for the qualitative evaluation of its faculty members. This gives them deeper insights about the quality standards of the teaching system as well as the teaching members. This data is very crucial for the managing bodies to take further steps and approaches to actively improve the very quality of teaching process and also make modifications if necessary.

The task of sentiment classification on a large scale is very tedious and time consuming, so many techniques have been developed over time to automate such work. Based on level of granularity sentiment analysis is generally classified as- Document level, sentence level and word level analysis. Many approaches simply calculate the polarity of a document or sentence by simply counting total no. of positive and negative words in the document or sentence. If the number of positive words is greater than that of negative then the overall polarity was calculated to be positive and vice-versa. However, we do not always use direct form of expression. Consider an example, *"the movie was definitely not bad."*, here the actual sentiment of this comment will certainly be misclassified if we account only for the number of positive and negative words. Hence, multiple machine learning and deep learning based models have been designed over time to increase the accuracy of sentiment classification.

There is another approach towards this task called the Aspect Based Sentiment Analysis. In this approach the context of the expressed sentiment is also considered along with the sentiment term for aspect wise classification. In our proposed model we are conducting Aspect Based Sentiment Analysis by using a deep learning based LSTM model which uses BERT for word embeddings. The preprocessed data is fed to the BERT embedding layer whose output is further passed on to the LSTM layers for aspect extraction and polarity classification. The task of aspect extraction is conducted by the LSTM layer1 and that of polarity classification by the LSTM layer 2.

II. LITERATURE SURVEY

A. S. Manek [1] used SVM binary classifier to find target aspects and opinion where they used semantic and various lexical features (unigram, bigram, and POS) and achieved an average accuracy of 94% on publicly available datasets. M. Sivakumar [2] calculated the semantic relatedness between opinion and aspect which finds the similarity between words using cosine similarity and uses traditional machine learning approach to find the polarity. This approach achieves precision of about 93.7%. However, this approach fails when the sentences contain "but" or "otherwise". I. Perikos [3] introduced ensemble classifier which includes Naïve Bayes, maximum entropy and support vector machines that are combined to get the sentiment of the users, this approach gave an accuracy of 87%. Y. Lv [4] proposed a four-layer system where he used BERT for word embedding which was then given to the aspect extraction layer that was further used to extract the summarized aspect representation from the contextualized sentence representation. Multi-granularity layer was then applied to capture the complex interactions between aspects and sentences. The approach out performed most of the available method for aspect-based opinion mining. It achieved an accuracy of 84.7% on publicly available restaurant dataset. I. Sindhu [5] proposed a domain specific model with LSTM layers and the domain word embedding done using word2vec that considers each word having only one vector. But in real world, depending on the context of the sentence, the meaning of the word may be different.

For example, “bank as a financial institute” and “bank of a river” are having meaning of the word ‘bank’ in different contexts. Due to this, aspect extraction is not that efficient with word2vec as embedding method. This model achieves the accuracy of 82% over a subset of SemEval-14. Z. Gao [6] proposed Target Dependent BERT(TD-BERT) where it takes the positioned output at the target words as input for classification. He found that both TD-BERT and BERT-FC(BERT-Vanilla) achieves slightly better performance than the neural network-based model. He also compared word embeddings with different models in which TD-BERT gives 78% accuracy on SemEval-2014 task4 dataset.

III. PROPOSED METHOD

The preprocessed data is fed to the BERT embedding layer whose output is further passed on to the LSTM layers for aspect extraction and polarity classification. The task of aspect extraction is conducted by the LSTM layer1 and that of polarity classification by the LSTM layer 2. The system architecture is depicted in Fig. 2.

BERT Embedding Layer

Generally, a word can appear multiple times in a single sentence having different meaning across different contexts. For example, “*the bank of Maharashtra and the bank of Yamuna”;* here the word ‘*bank*’ has different meaning in both contexts. The first ‘*bank’* refers to the financial institute and the second *‘bank’* is the shore of a river. The reason for using BERT is that it uses dynamic embedding instead of static embedding which considers context along with each distinct vector. Other techniques like Word2Vec assign a single vector for a particular token(word) whereas BERT assigns a distinct vector for each iteration of a token. In the above example Word2Vec assigns a single vector for the word ‘*bank’* which will be inefficient since it will not take the context of the word into consideration whereas BERT would assign two different vectors for the word ‘*bank’* for both the iterations along with their contexts.

[4] For a given sentence S, this layer maps a high-dimensional vector for each word in a sentence using pre-trained multilayer bidirectional Transformer blocks. Specifically, we obtain the input sequence S´ by concatenating the sentence with [CLS] and [SEP] and the final sentence S´ is obtained as [[CLS]; S; [SEP]], where [CLS] is a classification token and [SEP] is used as a sentence separator. For each token in the input sequence, the embedding layer assigns three different embeddings viz. token embedding, segment embedding and position embedding as given in Fig.1. Token embedding assigns a vector to each token in the input sequence while the segment embedding and position embedding layer differentiates sentence of token origination i.e. to which sentence the token belongs and the absolute position information respectively.

The input embedding is given as

$$S_0 = E_t[S'] + E_s + E_p \tag{1}$$

such that $S_0 \in R^{(n+2)*d}$ where,

S_0 = input embedding, E_t = token embedding, E_s = segment embedding, E_p = position embedding, n = length of sentence,

& d = number of hidden dimensions

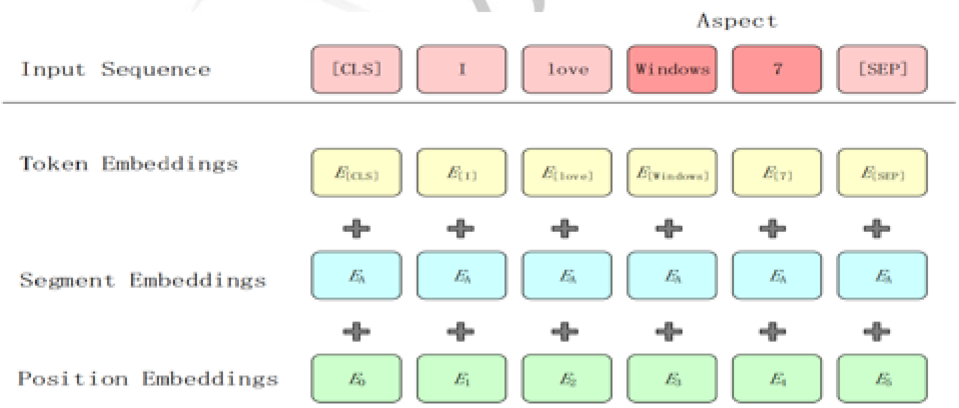

Fig.1 [4] BERT embedding layers.

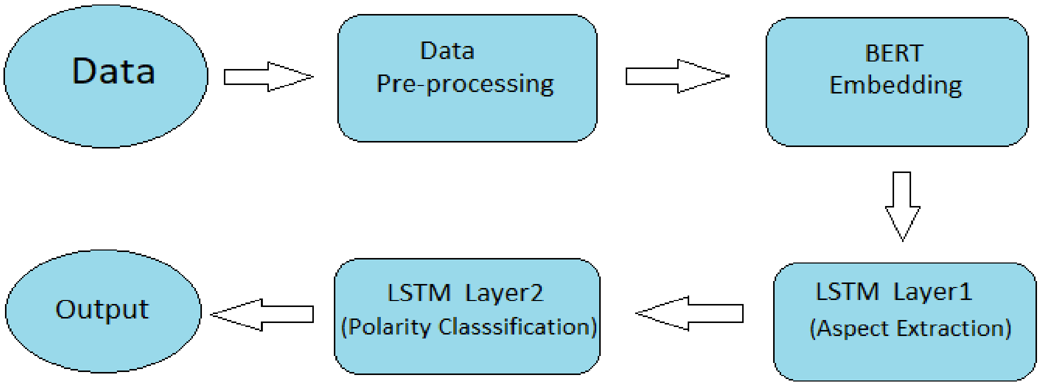

Fig.2 Proposed System Architecture

Aspect Extraction and Polarity Classification using LSTM

Here, we are going to use Long Short-Term Memory (LSTM) neural network for aspect extraction and polarity classification. LSTM is based on RNN but it overcomes the Long-term dependency problem as RNN learns from its immediate previous network. There may be cases where current input might not only depend on its immediate previous step, hence we chose LSTM over RNN.

[7] LSTM works on gating mechanism to control the information flow inside the LSTM cell as shown in Fig.3. LSTM has three gates namely forget gate, output gate and input gate which are denoted as *fg*, *og*, and *ig* respectively. The input embedding from BERT embedding layer $S0$ is passed through the forget gate of the LSTM cell which uses sigmoid function to decide whether to retain or discard the previous output.

$$fg = \sigma(Wfg.[ht\text{-}1, S0] + bfg) \tag{2}$$

where *fg* is the forget gate, *Wfg* is weight associated with forget gate and *bfg* is the bias for forget gate.
The input gate is responsible to decide what to add from current input into current cell with the help of sigmoid and tanh functions. The sigmoid function decides which part of current input should be retained and the tanh function generates the vector instance for the same.

$$ig = \sigma(Wig[ht\text{-}1, S0] + big) \tag{3}$$

$$ct' = \tanh(Wc[\,ht\text{-}1, S0] + bc) \tag{4}$$

where *Wig* is weight associated with the input gate and *big* is the bias.
The new cell state is obtained by concatenation as follows

$$ct = fg * ct\text{-}1 + ct' * ig \tag{5}$$

The current cell state *ct* is passed to the output gate which again uses sigmoid and tanh functions to decide what to output from the current cell state as output.

$$og = \sigma(Wog[ht\text{-}1, S0] + bog) \tag{6}$$

$$ht = og * \tanh(ct) \tag{7}$$

where *og* is output gate *Wog* is the weight associated with output gate and *ht* the current output.
Finally, softmax function is used as the final activation function which generates a probability distribution for all possible outcomes.

The second LSTM layer is used for polarity classification of the reviews. This layer takes the embeddings from BERT layer and the aspect prediction from the first layer. A probability distribution is created over three polarities – *negative, positive and neutral*. The polarity term with the highest probability is chosen as the final output or sentiment class.

IV. EXPECTED RESULTS

BERT based embedding is better than the available embedding methods like Word2Vec. Since, it provides dynamic embedding instead of static, it is capable of distinguishing between words used in different contexts. Moreover, we are going to use LSTM over RNN enabling long-term dependency. Our proposed system is expected to have higher accuracy over previous embedding methods and RNN based models.

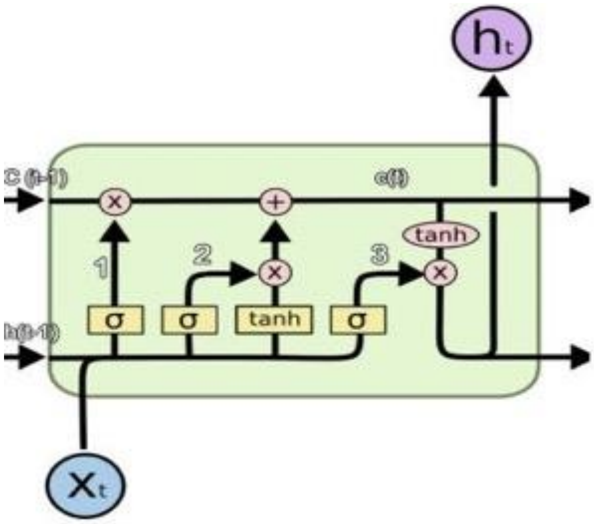

Fig.3 [8] Long Short Term Memory

V. CONCLUSION

Here, we propose aspect based sentiment analysis using BERT and LSTM which performs better than the available methods. Dynamic embedding in BERT is more powerful than static binding available in other methods like Word2Vec. Contexts of the sentences are better understood with the help of BERT. Aspect extraction and polarity classification is done using two layers of LSTM neural network.

VI. ACKNOWLEDGMENT

We would like to thank our guide Prof. Supriya Bhosale, Department of Information Technology, D. Y Patil College of Engineering, and my colleague Suraj Prajapati, Department of Computer Engineering for guidance and support.

REFERENCES

[1] A. S. Manek, P. D. Shenoy, M. C. Mohan, and V. K. R, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," World Wide Web, vol. 20, no. 2, pp. 135–154, 2016

[2] M. Sivakumar and U. S. Reddy, "Aspect based sentiment analysis of students opinion using machine learning techniques," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 726-731. doi: 10.1109/ICICI.2017.8365231

[3] I. Perikos and I. Hatzilygeroudis, "Aspect based sentiment analysis in social media with classifier ensembles," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017, pp. 273-278. doi: 10.1109/ICIS.2017.7960005

[4] Y. Lv, M. Hu, C. Yang, Y. Tang and H. Wang, "Extract, Attend, Predict: Aspect-Based Sentiment Analysis with Deep Self-Attention Network," 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Zhangjiajie, China, 2019, pp. 297-304. doi: 10.1109/HPCC/SmartCity/DSS.2019.00054

[5] I. Sindhu, S. Muhammad Daudpota, K. Badar, M. Bakhtyar, J. Baber and M. Nurunnabi, "Aspect-Based Opinion Mining on Student's Feedback for Faculty Teaching Performance Evaluation," in IEEE Access, vol. 7, pp. 108729-108741, 2019. doi: 10.1109/ACCESS.2019.2928872

[6] Z. Gao, A. Feng, X. Song and X. Wu, "Target-Dependent Sentiment Classification With BERT," in IEEE Access, vol. 7, pp. 154290-154299, 2019. doi: 10.1109/ACCESS.2019.2946594

[7] Y. Bengio, P. Simard and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," in *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157-166, March 1994. doi: 10.1109/72.279181

[8] https://towardsdatascience.com/understanding-lstm-and-its-quick-implementation-in-keras-for-sentiment-analysis- af410fd85b47